

Summary

Contents

A Quarterly Information Bulletin from the Methodology Division

Contents

Review of the Average Weekly Earnings Survey
Analysing Very Large Datasets
Developing Spatial Indexes
Impact of Large Weights on One Person Per Household Methodology
Quality Indicators for Household Surveys
Ken Foreman Award for 2000

Review of the Average Weekly Earnings Survey

REVIEW OF THE AVERAGE WEEKLY EARNINGS SURVEY

A reliable estimate of wages growth is an important economic indicator. One of the most popular series used by commentators is the Average Weekly Earnings (AWE) survey which produces the movement in full-time adult average weekly ordinary time earnings (AWOTE).

The quality of the published estimates has been questioned recently following the release of the August 2000 preliminary annual movement AWOTE figure of 5.9%. This movement was subsequently revised upwards in the final publication. Paul Sutcliffe and Bob McCombe were asked to review the survey, and report back to senior management on the quality of the published data. While the review had a wide ranging terms of reference this article concentrates on techniques for investigation of some of the methodological aspects.

The Frame Creation

The frame creation process has undergone a large number of changes in the past year and was therefore targeted for further investigation. Similarly the ABS's new business provisions (NBP's) were selected for review. NBP's represent the counts of businesses which did not make it on to our business register in time for selection. In order to manage these investigations, they were broken down into the following sections:

- A review of the population counts and NBP counts used by AWE;
- A detailed investigation into the creation of the Management Unit State (MUS) which is the statistical unit for AWE;
- A review of the process of handing over the frame creation and validation process to a central group within the ABS;
- Investigation into the effect of frame updates, especially benchmark employment and industry codes; and,
- A review of the calculation and application of the business provisions.

The review worked closely with the survey team to identify if there was anything unusual, or that had changed, which might have impacted on the published estimates.

The Decomposition of the Movement Estimates.

In order to better interpret the effects on the movement estimates of the changes in the survey frame and survey

sample over time, the decomposition components of the movement estimates can be collapsed into the following five key components:

- Common Sample Unit Effect;
- Net Birth-Death Effect = Frame Growth Effect + Sample Death Effect + Sample Birth Effect;
- Net Stratum Flip Effect = Frame Stratum Flip Net Effect + Sample Stratum Flip Out Effect + Sample Stratum Flip In Effect;
- Net Rotations Effect = Sample Rotation Out Effect + Sample Rotation In Effect Stratification Change Effect; and,
- Classification Change Effect.

Stratum flipping occur when information from a sample unit suggest that the unit should be in a different stratum. It is also possible to breakdown the common sample unit effect into different categories. Each unit's contribution to the categories can be ranked. In this way any unusual impacts can be readily identified.

A macro to evaluate these impacts is now available. This tool proved valuable in completing the AWE review as it allowed us to evaluate the impact of frame changes on the estimates. It also provided lists of units having high impacts on the movements which could then be identified and interrogated.

The Direct Movement Estimator.

The ABS currently designs the AWE survey to produce optimal level estimates of AWOTE. Due to the existence of highly correlated quarterly samples (in general) and the strong correlation between the data items collected from the same business, the quarterly differences of the level estimates are a good measure of the quarterly movements in AWOTE. The annual differences in the level estimates are not as robust as the quarterly as there is not as much common sample and changes in the composition of the population may affect the results. A direct movement estimator is a more robust measure for both quarterly and annual movement, as it is based on the common sample between the two time periods and allows for population changes between them.

The ABS has always been active in reviewing methodologies for keeping the frame as up to date as possible and has presented papers to conferences on this topic. The use of decomposition techniques to understand the contribution to estimates of different components allows the survey teams to fully understand their data. The direct movement estimator is used by a number of survey teams to review the published data and understand the main drivers behind the estimates.

For more information, please contact Paul Sutcliffe on (02) 6252 6759.

E-mail: p.sutcliffe@abs.gov.au

Analysing Very Large Datasets

ANALYSING VERY LARGE DATASETS

Government agencies and businesses are accumulating large databanks that potentially have considerable value for statistical purposes. The ABS has initiated projects to explore and exploit that potential:

- We are negotiating protocols with government agencies to set standards for the storage, vetting and documentation of administrative data.
- We are considering the implications of electronic commerce and business-to-business data exchange for our future data capture and estimation systems.
- We are exploring the statistical application of particular datasets, such as the Australian Taxation Office's Business Activity Statement and the scanner data collected at supermarket checkouts.

Over the years, Methodology Division and other parts of the ABS have developed a large array of mathematical tools and computer software to help us analyse datasets collected through the bureau's own censuses and sample surveys. The question arises, however, whether those tools and software will remain appropriate when we must deal with very large "by-product" datasets.

- Can the statistical analysis software we use today cope with very large (and rapidly growing) datasets? If not,

what new computing tools do we need to store, transport, browse and transform these datasets?

- How might our traditional models have to be changed to deal with datasets that have not been assembled using ABS classifications, variable definitions and collection methods? What statistical methods do we need to navigate and manipulate these datasets? What methods do we need to assess their quality?
- How might our research strategies have to change? Might we do the bulk of our exploratory analyses on sampled datasets, then validate our preferred or final model against the full dataset? Can emerging techniques for data mining and data visualisation help us?

During 2000 and 2001, a group of staff from Methodology Division, ESG Strategic Development Section and Technology Applications Branch are thinking about these questions. We shall be contacting other statistical agencies and owners of large datasets, reviewing recent software developments, and ransacking the statistical literature for techniques and tools that look promising.

For more information, please contact Ken Tallis on (02) 6252 7290.

E-mail: ken.tallis@abs.gov.au

Developing Spatial Indexes

DEVELOPING SPATIAL INDEXES

There is considerable interest in comparing price levels between different parts of Australia. Policy designers would like to know the relative costs of living in different places. Businesses would like to know what allowances to pay when employees move from one city to another. Relative price levels appear in the models used by economists to analyse questions about the locational choices of industries and households.

The Consumer Price Index (CPI) tells us about the rates at which prices are changing in the eight capital cities; it does not tell us about relative price levels or relative costs of living. The ABS Analysis Branch, Prices Development Section and the Consumer Price Indexes Section are developing some experimental "spatial indexes", with the aim of filling this gap.

During the first half of 2001, the project team will be undertaking some exploratory work to address such questions as:

- What questions is a spatial index attempting to answer?
- What goods and services should be or can be included in a spatial price analysis?
- What index formulae are most appropriate?

The ABS has done some work of this kind in the past. For example, in 1994 there was a three-way comparison of the cost of living in Darwin, Perth and Sydney, to support the inquiries of a government committee. In 1995, a comparison was done for Christmas Island, the Cocos Islands and Perth. Our exploratory work next year will draw on that experience.

One of the technical problems to be solved arises because many items included in the CPI basket for one city may not appear in the baskets for other cities. Ways of dealing with this "sparse matrix" problem have been proposed by several analysts, including Dr Jim Cuthbert (former head of the Scottish National Statistical Office) who visited the ABS in early 2000. The project team will be consulting Dr Cuthbert on this and other issues.

For more information, please contact William Milne on (02) 6252 6298.

E-mail: william.milne@abs.gov.au

Impact of Large Weights on One Person Per Household Methodology

IMPACT OF LARGE WEIGHTS ON ONE PERSON PER HOUSEHOLD METHODOLOGY

A number of collections within the ABS household survey program enumerate one individual within selected households. The initial weight is calculated as the number of in-scope persons in the household, multiplied by the household selection weight. This weight is then post stratified. Thus random persons in large households can have relatively large weights and, if they have a relatively low or high value for particular data items, could "unduly" influence the estimate for those data items.

The usual method for addressing this situation would be some form of outlier technique but there may be better ways to adjust for this which does not involve "tampering" with what is quite possibly valid data. As the problem of unstable estimates arises from the selection method rather than data outliers a better approach might be to address the issue of weight outliers.

One approach to reducing the impact of extreme weights is to select more persons from those households which would likely generate the large weights. This can be achieved by selecting two random persons from large households thereby effectively halving the weight and subsequently halving the impact of selected persons with extreme, but valid data values and large weights.

An annual file from the 1998/99 Population Survey Monitor was analysed to examine the impact of outliers on estimates. This showed there were 127 households that had greater than five persons per household, and hence were classified as 'large households'. Weight analysis showed that the weights increase as household size increases. However the size of the increase is not as large as would be expected for the maximum weights. This suggests that the post-stratification also has a large influence on the size of the weight.

A sample was synthesised to reflect the selection of two persons from large households by randomly selecting additional persons from the 127 households and adjusting the weights accordingly. Income data was imputed for the new selections and estimates of deciles were derived. The effect on the deciles was small, and the differences were limited to only the 4th, 6th and last deciles. Even then, the differences are minor, with the greatest change in decile cutoffs being \$2000 (annual income).

Looking more closely at the weights showed that what made the weight large had more to do with the demographic variables and area that the respondent was from, than the size of the household. This is reflected in the maximum and minimum weight analysis, where the weights for respondents in households of 7 people range from 357 to 6,678.

The results of this investigation do not suggest that we need to be overly concerned about the impact of large weights arising due to the selection of one person from large households. However, it remains a possibility that a combination of large weights and extreme values could "unduly" influence estimates.

For further information please contact Julie Watkinson on (08) 8237 7539

Email: julie.watkinson@abs.gov.au

Quality Indicators for Household Surveys

QUALITY INDICATORS FOR HOUSEHOLD SURVEYS

The ABS is currently developing a standard set of quality indicators relevant to most household surveys run by the ABS. These quality indicators cover all components of the survey cycle, and together they provide a consolidated framework of indicators from which users can assess the quality of survey data and ABS can monitor quality issues. Data quality encompasses many aspects, especially accuracy, timeliness, relevance and cost of production.

The quality indicators will be implemented in three phases, based on the importance of the indicator and the extent of work required to implement it. Phase 1 indicators involve those indicators that are relatively easy to implement in terms of data availability and cost. Phases 2 and 3 indicators require greater effort in terms of developing methodology, setting up procedures and resolving cost implications.

Many facets of the survey cycle are covered in the core set of quality indicators, from aspects such as frame quality and sample design, through to the questionnaire and timeliness. However, accurately measuring the quality of some aspects, especially the questionnaire, is difficult to do quantitatively and hence these quality indicators should be used with caution. Although many of the quality indicators will only be used internally within the ABS, some of them may be useful to external clients. Both internal and external users of ABS household survey data should be able to use, and should actually be using, information about the quality of our data. This information should be used to help make appropriate judgements on the quality of the data when making informed decisions.

Quality indicators for household surveys are useful for:

- Assisting survey managers and data users to assess the quality of survey data and determine whether that quality is sufficiently good for the intended use of survey outputs.

- Providing an objective basis for continuous improvement of the survey process.
- Enabling managers to optimise the resources spent on different parts of the survey process.

Monitoring of appropriate quality indicators provides a mechanism for ensuring that survey outputs are of an appropriate quality for their intended use. In addition to routine monitoring of quality indicators, survey managers should conduct special investigations aimed at identifying and reducing/removing the factors which would (or would likely) cause a reduction of quality. For continuous improvement of the survey operation, it is important that quality indicators on components of a survey operation be evaluated on a regular basis.

It is envisaged that in future a standard set of quality indicators will be produced for each ABS household survey. By standardising the measures to be produced, their definitions and the infrastructure for their production and dissemination, both internal and external users will be able to compare the quality of data between different surveys, as well as monitor, for repeated surveys, the change in quality over time. In addition, survey areas will be able to adopt best practices in quality management and help raise ABS's capability in the measurement and dissemination of quality information.

In addition to the core set of standard quality indicators that encompass all surveys, there are other important quality indicators which are specific to the content or methodology of each survey which will be implemented on a survey basis.

This project is one of several ABS projects that are currently investigating the issue of quality.

For further information, please contact Matthew Cronin on (02) 6252 5594.

Email: matthew.cronin@abs.gov.au

Ken Foreman Award for 2000

KEN FOREMAN AWARD FOR 2000

In commemoration of the contribution made by Mr Ken Foreman to the ABS over many years, the Australian Statistician agreed to institute an annual award for an officer of the Methodology Division who is performing at a high level, and is showing the potential for substantial further development as a methodologist in the ABS, and the ability to make a significant contribution within the ABS. The award comprises an overseas trip to an international conference or short training course.

The 2000 Ken Foreman Award has been awarded to Kristen Northwood of the Analytical Services Branch. In receiving the award, it is proposed that Kristen attend the 2001 Siena Summer School on Social Statistics, visit Statistics Netherlands with a view to discussing their System of Economic and Social Accounting Matrices and Extensions and visit both Statistics Canada and the United States Bureau of the Census to discuss related analytical projects.

Kristen graduated from the Australian National University with Bachelor of Economics and Bachelor of Science degrees. She first joined the ABS in 1996, and joined Methodology Division in September 1998, having transferred from the Census Evaluation area in Population Surveys Group. She worked in Analytical Services until the end of 1999, her major contributions being on the longitudinal analysis of the Growth and Performance Survey (GAPS) data set, and the creation of experimental output measures for the Justice sector. She then joined the newly-created Analysis Branch in January 2000, where she has completed the Justice output measurement project, and commenced work on extending and refining estimates of Australian household wealth.

Previous recipients of the award have been:

- 1996 Robert Clark - Strong technical contribution in his work as demonstrated in the Labour Force Survey redesign.
- 1997 Steven Kennedy - Demonstration of good technical knowledge and skills, as well as the willingness to further develop his skills through formal study, and in the course of his work. Part of a team responsible for undertaking a significant part of the productivity analysis project. Has represented the ABS and its methodological interests and helped build some useful networks with others who were conducting research in the area.
- 1998 Dina Neiger - Contribution to developing the methodology for business collections, particularly through the review of the sample and frame maintenance procedures (SFMP).
- 1999 Richard McKenzie - Strong contribution to the development of the methods used in the Labour Cost

Index and playing a lead role in the development and implementation of a novel 'rotating panel' method of selecting the sample for the Labour Cost Index. Has also worked on significance editing and significance follow up.

For further information on the Ken Foreman Award, please contact Bill Allen on (02) 6252 6302.

E-mail: bill.allen@abs.gov.au

About this release

Methodological News

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence